DOCUMENT RESUME

ED 405 216 SE 059 805

AUTHOR Brewer, Steven D.
TITLE A Model of Desired Performance in Phylogenetic Tree Construction for Teaching Evolution.
PUB DATE Mar 97
NOTE 27p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (70th, Oak Brook, IL, March 21-24, 1997).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Biology; *Concept Formation; Educational Strategies; *Evolution; Higher Education; Learning Strategies; Models; *Problem Solving; Science Instruction; Secondary Education

ABSTRACT

This research paper examines phylogenetic tree construction-a form of problem solving in biology-by studying the strategies and heuristics used by experts. One result of the research is the development of a model of desired performance for phylogenetic tree construction. A detailed description of the model and the sample problems which illustrate each step are included. The study involved expert phylogenetic systematists (N=9) who used think-aloud protocols as they problem solved. A discussion of the use of the model which was developed during this study is also included. It points out that the model can be useful to students to help them interpret the instructor's behavior and to guide their own problem solving. The model by itself would not be sufficient to teach effective problem solving to students. The tree construction problems used in the study allow students to become familiar with the processes used by scientists to explain evolutionary history. (DDR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Running Head: DESIRED PERFORMANCE IN TREE CONSTRUCTION

A Model of Desired Performance in Phylogenetic Tree Construction for Teaching Evolution

Steven D. Brewer

Western Michigan University

ED 405 216

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S D Brewer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Abstract

Although evolution is the central theory in biology, many students fail to understand it and leave biology instruction unable to appreciate its importance. A problem-based approach to teaching and learning evolutionary biology may offer a number of benefits to students. One form of problem-solving in evolutionary biology, phylogenetic tree construction, was examined by studying the stratgies and heuristics used by experts. One result of this research has been a model of desired performance for phylogenetic tree construction. This report describes this model in detail and illustrates each step using an example problem that was challenging to experts. Used in an appropriate instructional setting, this model can result in good student problem solving in phylogenetic tree construction.

A Model of Desired Performance in Phylogenetic Tree Construction for Teaching Evolution

Evolution is undoubtedly the most important theoretical framework in biology. Unfortunately evolution is rarely accorded a place in the biology curriculum commensurate with its importance within biology theory. Evolution is often simply equated with natural selection and taught from a primarily functional perspective. Comparative and historical approaches, that are critical for developing an appreciation of the power of evolutionary theory, are often neglected. This contributes to evolution being poorly understood and widely disparaged among both teachers and American society at large. This paper describes results from a research program situated within the problem-solving tradition in science education to improve the teaching of evolution.

A problem-based approach to the teaching and learning of evolution may offer a number of benefits to students. Stewart (1988) has outlined four classes of potential learning outcomes from the use of problem-solving in genetics: (a) the conceptual structure (laws, theories, and their organization) of a particular discipline; (b) problem-solving heuristics that are not specific to a particular discipline; (c) content-specific problem-solving procedures (domain-specific instantiations of general heuristics and problem-solving algorithms); and (d) insight into the nature of science as an intellectual activity. Similar potential learning outcomes are likely from a problem-based approach to the teaching of evolution.

An approach to teaching science developed by the BioQUEST Curriculum Consortium offers greater potential learning outcomes for students than other more traditional approaches (Jungck & Calley, 1985). This approach has been called the "3 P's": problem posing, problem solving, and peer persuasion. To implement this approach successfully, however, a more extensive knowledge base is required than for traditional instruction (Reif, 1983). Teachers must be familiar not only with the conceptual

knowledge of a domain, but also the strategic knowledge necessary to engage in effective problem-solving. In addition to solving problems, however, teachers must also have this knowledge organized in a form that will facilitate instruction. To be successful, instruction in solving problems requires a knowledge base composed of at least three bodies of information: (1) conceptual structure that relates tasks to conceptual knowledge, (2) relevant problems that encompass the range of phenomena to be addressed, and (3) explicit procedures that include: (a) models of problem solving that can lead to success and (b) strategies and heuristics that can guide how to implement those models across the full range of situations that students may encounter.

Although genetics problem solving and instruction has been relatively well studied from this perspective (See Stewart and Hafner 1994 for a review), most other areas have not. This is especially true of areas that have not traditionally been conceptualized from a problem-solving perspective, like evolution. This report represents an initial attempt to apply techniques from the problem solving research tradition in science education to the domain of evolutionary biology. A research project carried out during 1995 reviewed the literature on phylogenetic biology and studied experts constructing phylogenetic trees. This report describes the model of desired performance in problem-solving that was developed based on this research.

Methods

An initial literature review provided insight into basic phylogenetic problems and methods. Among others, Ridley (1986) and Brooks & McLennan (1991) provided an overview; Eldredge and Cracraft (1980) and Wiley (1981) provided insight into the nature of phylogenetic problems and solutions; and Wiley, Siegal-Causey, Brooks, & Funk (1991) provided a primer of methods. The literature review resulted in: (a) a statement that illustrates the situations in which phylogenetic inference is useful, (b) a statement that relates tasks to conceptual knowledge, and (c) the development of

Phylogenetic Investigator, a software problem-solving environment that was used to present problems to experts (Brewer and Hafner, 1996).

A problem-solving research methodology was developed based on Larkin and Rainard (1984), Ericsson and Simon (1993), and Ericsson and Smith (1991). Two types of phylogenetic problems were constructed. The first type, termed "model problems", represented an attempt to generate a typology of the problematic phenomena in phylogenetic tree construction. These model problems are incorporated with the Phylogenetic Investigator program. The second type of problems, termed "research problems", represented an attempt to create a range across factors that lead to difficulty in phylogenetic problems. Three series of research problems were constructed that varied the numbers of solutions, taxa, and characters. Each problem consisted of a matrix of coded and polarized phylogenetic data organized by taxa and characters. In addition, a fourth series of problems contained revision components that required additions to prior solutions, restructuring of prior solutions, or increased or decreased numbers of solutions.

Nine expert phylogenetic systematists participated in the research project by thinking aloud while constructing phylogenetic trees to account for the problem data matrices. The think-aloud protocols and the recorded actions from the problem-solving environment were collected along with all notes and drawings. These data were used to develop a descriptive procedural model of expert performance for phylogenetic tree construction. A synthesis of the model of expert performance and original analysis of phylogenetic problems was used to develop a model of desired student performance in phylogenetic inference. The model provides a basis for developing an approach to teaching evolution based on effects-to-causes problems.

Results

Experts used three overall strategies to construct phylogenetic trees. Based on the most commonly used strategy and heuristics from several different experts, a model of desired performance (Fig. 1) was constructed that presents a prescriptive series of

procedures for phylogenetic tree construction. The procedures synthesize components from the model of expert performance into a set of steps that can be incorporated into teaching practice. This model should lead to good problem solving in phylogenetic tree construction and be applicable either to computer-based or paper-and-pencil approaches to phylogenetic tree construction. Below the model is described in detail and is used to work through an example problem. The model description and example problem both assume familiarity with the fundamental concepts and ideas in phylogenetic biology. Readers are encouraged to refer to Appendix A which contains a primer of phylogenetic biology as an introduction and reference.

(1) Organize the characters, mentally, in the matrix, or on paper, to find the largest inclusion/exclusion character group hypothesis or hypotheses. (Consider the order of the taxa in the matrix as a mechanism of enhancing inclusion/exclusion patterns).

(2) Translate a hypothesis into taxa by organizing the taxa in the drawing field.

(3) Postulate an ancestor for each character or group of identical characters in the inclusion/exclusion hypothesis.

(4) Link the most inclusive ancestors, to the next less inclusive ancestors, and continue until reaching the terminal taxa.

(5) Distribute homoplasious characters.

(6) Considering each homoplasious character, starting with the character that requires the most steps:

  (6a) Evaluate whether options that improve its distribution always result in matching losses in other characters

  (6b) Evaluate whether other homoplasious characters have similar distributions that reinforce each other

(7) Construct other topologies based on additional inclusion/exclusion hypotheses from (1) or parsimony hypothesis from (6a) or (6b)

8) For each topology consider alternate optimizations for each homoplasious character.

Figure 1. A model of desired performance for phylogenetic tree construction.

The example is based on one of the research problems that was provided to the experts to solve. This problem, problem 1.4, proved to be very challenging to experts -- no expert solver discovered all of the three most parsimonious topologies, seven of nine experts found two, one expert found one, and one expert found none of the most parsimonious topologies. Figure 2 illustrates the form in which the problem was presented.

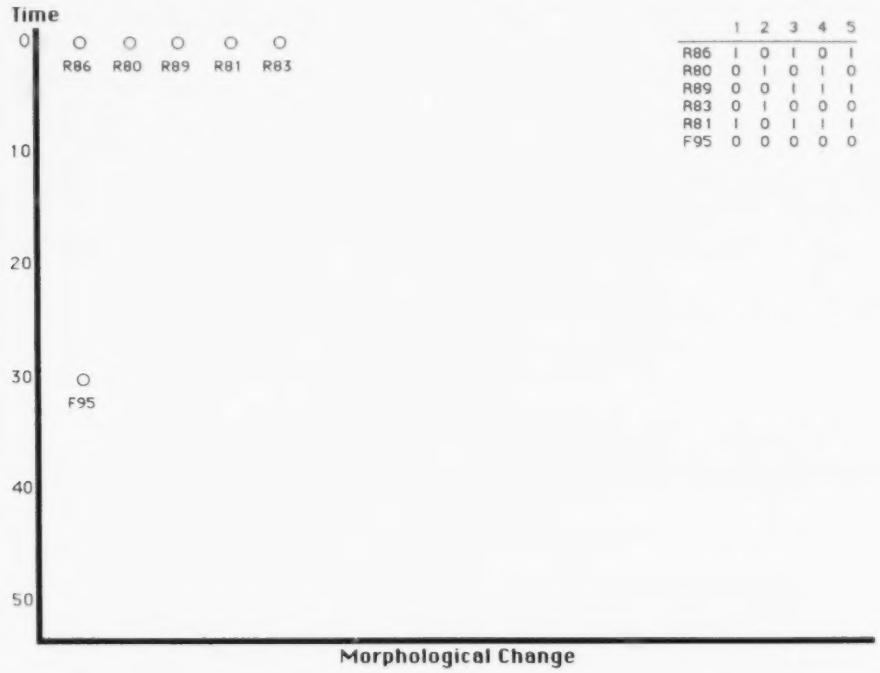

Figure 2. Problem 1.4 as it was presented to the problem solver. The taxa are arranged in a random order in the drawing field. The data matrix, presented at the upper right in this figure, is actually displayed in a separate window when using Phylogenetic Investigator.

(Step 1) Organize the characters, mentally, in the matrix, or on paper, to find the largest inclusion/exclusion character group hypothesis or hypotheses. (Consider the order of the taxa in the matrix as a mechanism of enhancing inclusion/exclusion patterns).

Inclusion/exclusion character group hypotheses represent groups of characters that tell a consistent story. Usually one starts with the character with the largest number of taxa in the apomorphic state, and one looks for other characters that have a distribution of

taxa in the apomorphic state that is inclusive (a subset) or exclusive (contains no overlapping members) of the first character. These sets of characters are used as the basis for constructing a phylogenetic tree. The largest set of such inclusive/exclusive characters does not always lead to a most parsimonious solution, but usually leads to a very good first approximation.

In the example problem, Characters 3 and 5 are identical, and character 2 is exclusive from them. Character 1 is inclusive with respect to characters 3 and 5. Character 4 conflicts with all other characters. The largest inclusion/exclusion group hypothesis is {1,2,3,5} which will serve as the basis for an initial solution.

Organizing the characters in the matrix can be a useful heuristic for finding inclusion/exclusion hypotheses. One heuristic is to count the number of apomorphies for each taxon and to order them in the matrix -- the taxa with the most at the top. Moving similar taxa and characters together can emphasize the pattern of inclusion/exclusion.

In the example problem, taxon R81 has four apomorphies and is moved to the top. Taxon R86 has three, R89 has three, R80 has two, and R83 has one. Characters 3 and 5 are identical and are placed adjacent to each other. The inclusive relationship between 1, 3, and 5, is emphasized by moving character 2. Character 4 conflicts with all other characters and is moved all the way to the right. Figure 3 contrasts the data matrix in its original form, after being ordered, and after being organized.

	1	2	3	4	5
R86	1	0	1	0	1
R80	0	1	0	1	0
R89	0	0	1	1	1
R83	0	1	0	0	0
R81	1	0	1	1	1
F95	0	0	0	0	0

	1	2	3	4	5
R81	1	0	1	1	1
R86	1	0	1	0	1
R89	0	0	1	1	1
R80	0	1	0	1	0
R83	0	1	0	0	0
F95	0	0	0	0	0

	1	3	5	2	4
R81	1	1	1	0	1
R86	1	1	1	0	0
R89	0	1	1	0	1
R80	0	0	0	1	1
R83	0	0	0	1	0
F95	0	0	0	0	0

Figure 3. The matrix on the left represents the random initial presentation of the data in the problem. In the middle, the taxa have been ordered by number of apomorphies. On the right, the characters have been organized to emphasize the inclusion/exclusion relationships among characters.

(Step 2) ***Translate a hypothesis into taxa by organizing the taxa in the drawing field. Taxa that share characters, and groups of taxa representing monophyletic taxa that share characters should be placed in adjacent locations on the screen.***

Each character or group of identical characters in the inclusion/exclusion hypothesis will be used to define a monophyletic taxon, represented by a common ancestor that will be shared by a group of descendant taxa. The smallest groupings will be subordinate to the larger groupings leading eventually to a single common ancestor.

Using the inclusion/exclusion group hypothesis of {1,2,3,5}, taxa R86 and R81 are placed adjacent to each other on the basis of character 1. Taxa R80 and R83 are placed adjacent to each other on the basis of character 2. Taxon R89 and the putative monophyletic taxon of R86 and R81 are placed adjacent to each other. All of the taxa are assumed to be descendants of taxon F95. Figure 4 illustrates the drawing field after the taxa have been organized.

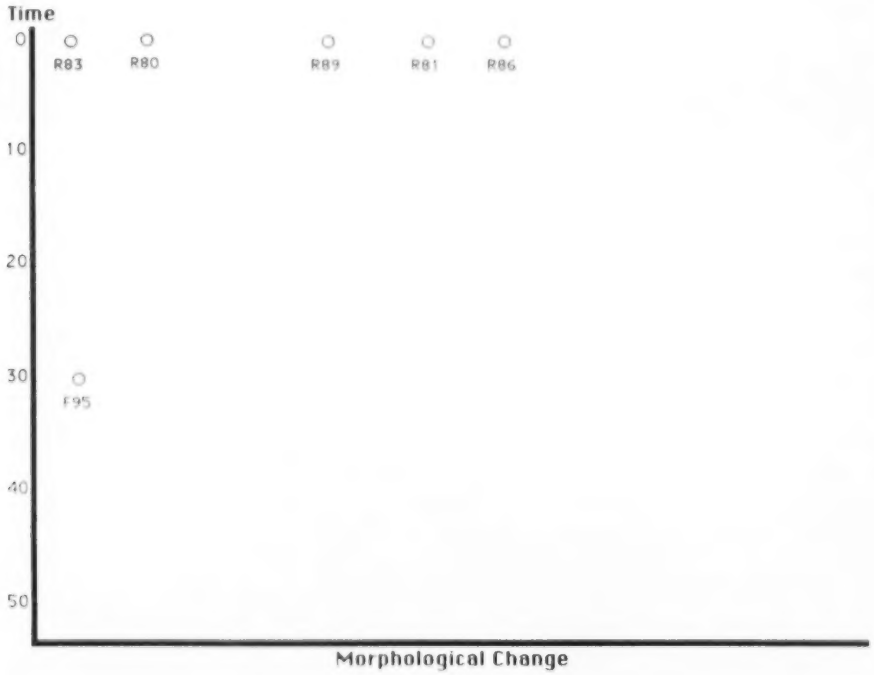

Figure 4. The taxa have been organized in the drawing field to represent the relationships inferred from the initial inclusion/exclusion hypothesis.

(Step 3) Postulate an ancestor for each character or group of identical characters in the inclusion/exclusion hypothesis.

The postulated ancestors represent the branching points prior to which characters shared in the apomorphic state must have changed from plesiomorphic to apomorphic. Postulated Ancestor PA is added for character 1, PB for character 2, and PC for characters 3 and 5. Figure 6 illustrates the drawing field after the postulated ancestors have been added.

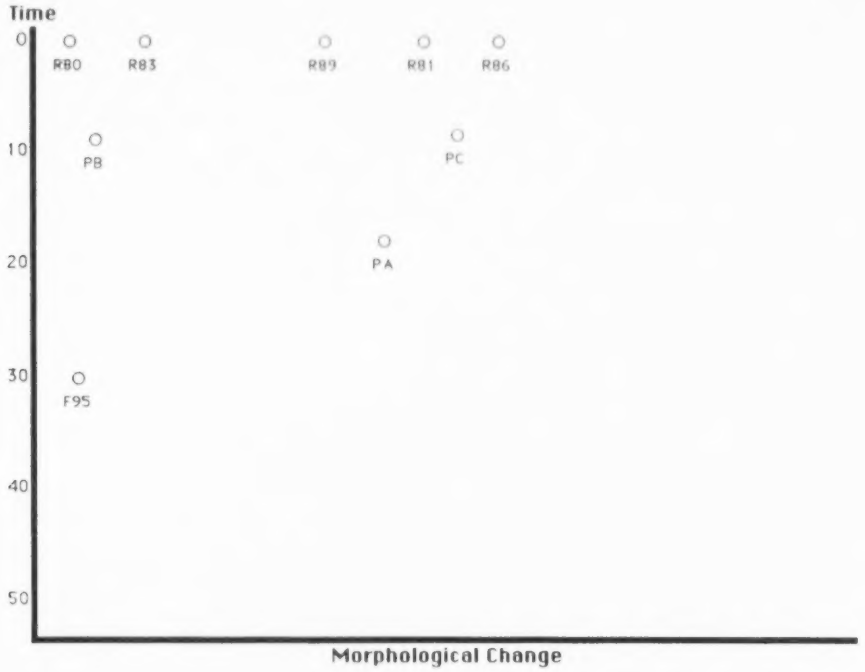

Figure 5. Problem 1.4 after postulated ancestors have been added.

(Step 4) Link the most inclusive ancestors, to the next less inclusive ancestors, and continue until reaching the terminal taxa.

In constructing the initial tree, homoplasious characters (in this case, only character 4) are ignored. Taxon F98 is the most inclusive ancestor -- all of the taxa are assumed to be descended from F98. Three taxa, R84, R87, and R85 are inferred descendants of PC based on the evidence of characters 3 and 5. A link is formed between F98 and PC. The other two taxa, R83 and R81, are inferred ancestors of PB based on the

evidence of character 2. A link is formed between F98 and PB. Taxon PB is now identical in state with R81 and R83 and links can be formed between them. Taxon R89 is identical in state with PA and a link can be formed between them. A link can be formed between PA and PC on the basis of character 1. Taxa R81 and R86 are then identical in state with PC and can be linked. Figure 6 illustrates the initial tree.

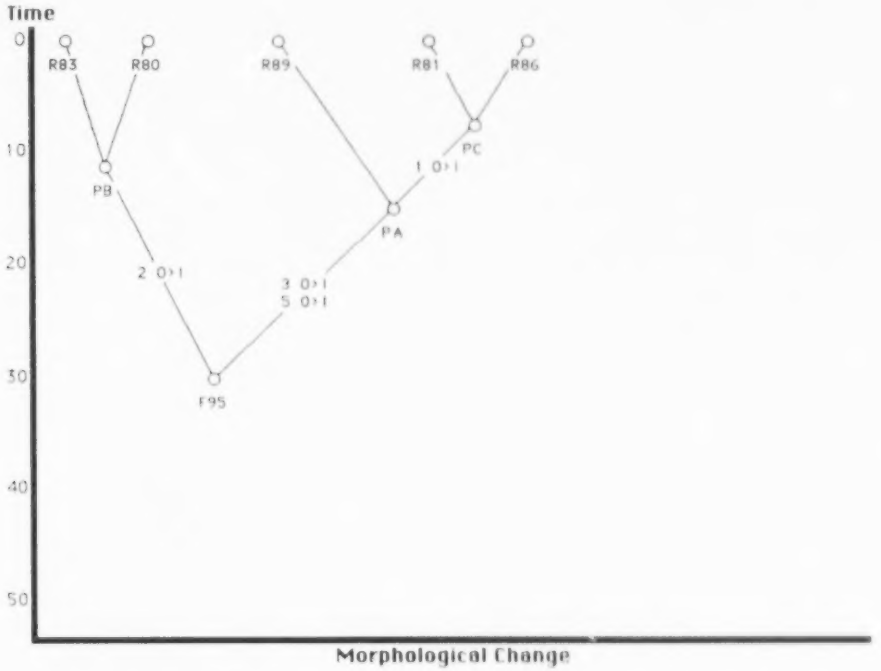

Figure 6. The phylogenetic tree with only the non-homoplasious characters of the largest inclusion/exclusion hypothesis distributed.

(Step 5) Distribute homoplasious characters.

Homoplasious characters, typically represent evidence for different possible reconstructions. In constructing the initial tree, the evidence they represent is explained as multiple changes in character state. Initially, it is desirable to explain these characters as multiple forward transitions. In the last step, the potential for these characters to be explained as reversals is explored.

Character 4, the only homoplasious character using this inclusion/exclusion hypothesis, can be distributed as three forward transitions prior to taxa R80, R89 and

R81. Figure 7 illustrates the initial solution to this problem constructed based on the largest inclusion/exclusion hypothesis.

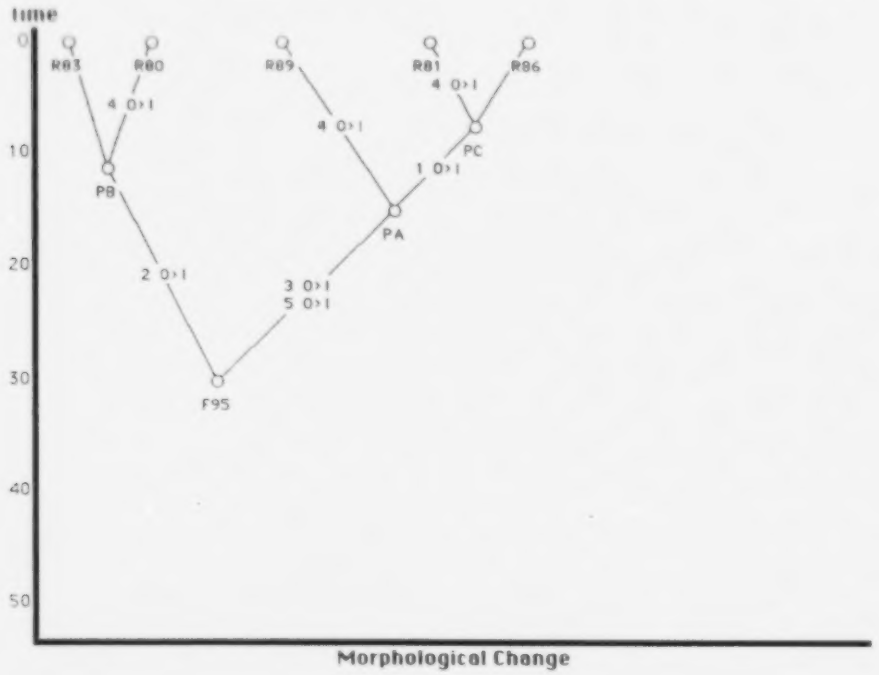

Figure 7. A most parsimonious solution to problem 1.4 based on the largest inclusion/exclusion hypothesis {1,2,3,5}.

(Step 6) Considering each homoplasious character, starting with the character that requires the most steps:

(6a) Evaluate whether options that improve its distribution always result in matching losses in other characters

(6b) Evaluate whether other homoplasious characters have similar distributions that reinforce each other

At this point, the goal is to consider the evidence provided by homoplasious characters and see if this evidence can be understood as alternate reconstructions of evolutionary history that do not require more transition events. The criterion of using the number of transition events, termed "parsimony", is controversial among scientists and philosophers. It is probably the most widely adopted criterion in use today for phylogenetic biology.

Character 4 is the only homoplasious character. It requires 3 steps given the current hypothesis. One evaluates options that improve the distribution of a character by asking how other characters would be affected if a tree were constructed in which the homoplasious character was not homoplasious (or was less homoplasious). There are no ways to rearrange the tree that will reduce the number of transitions required to explain character 4 without also affecting the number of steps required to explain other characters. There are several ways, however, to rearrange the taxa and save one or more steps from character 4. Most of these ways require an increase of more than the same number of steps in another character or characters. There are two hypotheses that improve character 4 by one step that result in only matching increases in other characters. These represent equally parsimonious solutions of the problem. Each of these other equally parsimonious hypotheses is based on a smaller inclusion/exclusion hypothesis: {2,3,5} or {1,3,5}.

(Step 7) Construct other topologies based on additional inclusion/exclusion hypotheses from (1) or parsimony hypothesis from (6a) or (6b)

In the first alternate hypothesis, we can reduce the number of steps required to explain character 4 by increasing by one the number of steps to explain character 1. In this hypothesis (Fig. 8) a monophyletic taxon of R81 and R89 share a common ancestor PC supported by the evidence of character 4. Character 1 is now explained as two forward transitions prior taxa R81 and R86.

The final equally parsimonious topology (Fig. 9) explains character 2 as separate forward transitions prior to taxa R80 and R83. A forward transition for character 4 supports a common ancestor for a monophyletic taxon composed of R80, R89, R81, and R86. Character 4 also requires a backward transition (from 1 to 0) prior to taxon R86. Characters 3, 5, and 1 are explained as in the first topology (Fig. 7).

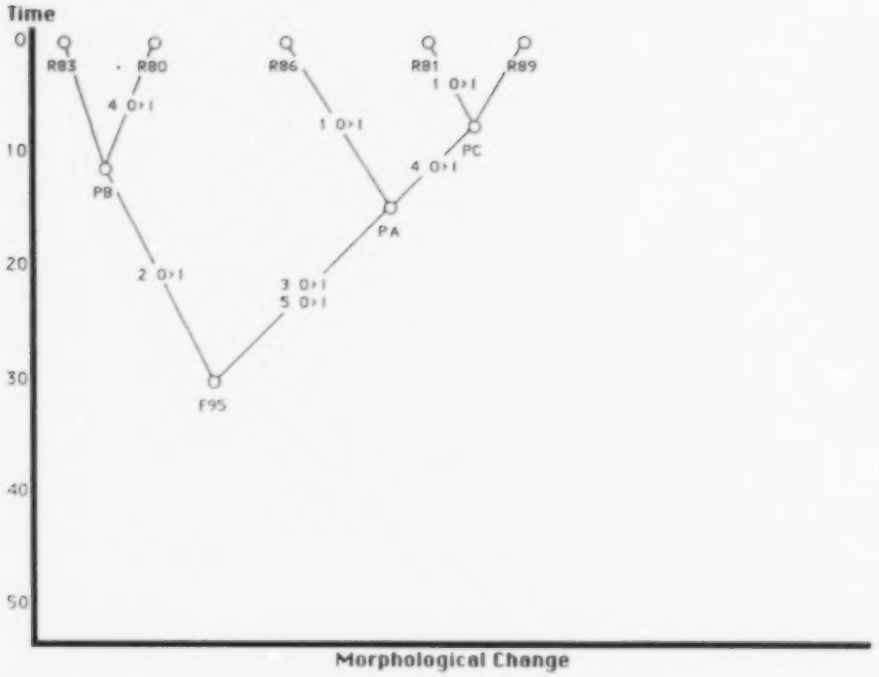

Figure 8. A second most parsimonious topology for problem 1.4 based on the inclusion/exclusion hypothesis {2,3,5}.

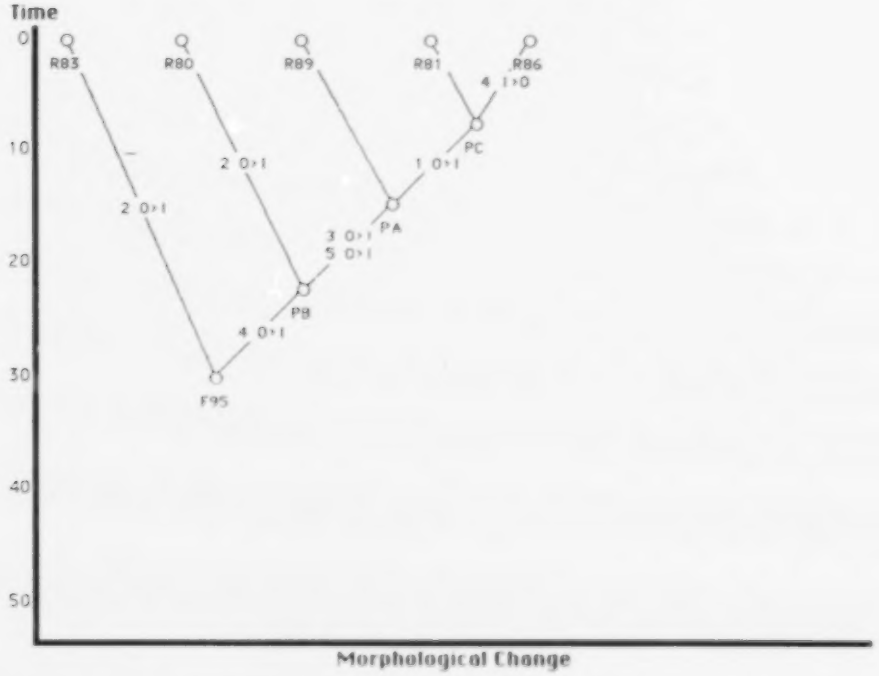

Figure 9. A third most-parsimonious topology for problem 1.4 based on the inclusion/exclusion hypothesis {1,3,5}.

(Step 8) For each topology consider alternate optimizations for each homoplasious character.

In most cases that there are multiple forward transitions for a character, equally parsimonious reconstructions exist that combine a single forward transition and a subsequent reversal or reversals. Starting with the first topology, character 4 can be optimized three ways: as pictured in Figure 7 with three forward transitions (prior to taxa R80, R89, and R81; not pictured with two forward transitions (prior to taxa R80 and PA) with one reversal (prior to R86), and as pictured in Figure 10 with one forward transition (prior to taxon PE) with two reversals (prior to taxa R83 and R86). The second topology has two character optimizations : as pictured in Figure 7 with two forward transitions (prior to taxa R86 and R89) or one forward transition (prior to taxon PA) and one reversal (prior to taxon R89). The third topology also has two character optimizations: as pictured in Figure 8 with two forward transitions for character 2 (prior to taxa R83 and R80) and as pictured in Figure 11 with a forward transition for character 2 prior to PE and a reversal prior to PA.

This problem, in spite of having only a handful of characters and taxa, is quite difficult to solve completely, even for experts. A complete solution to problem 1.4 requires recognition of three topologies, with the three optimizations of topology 1, and two optimizations for each of the other two topologies. Although some topologies are quite easy to find, others are more difficult. The model of desired performance presented here, in conjunction with good instruction and practice, has the potential to allow students to completely solve problems of this complexity. In the process, the fundamental concepts of phylogenetic biology become familiar to students and they develop a new way of perceiving evolutionary history.

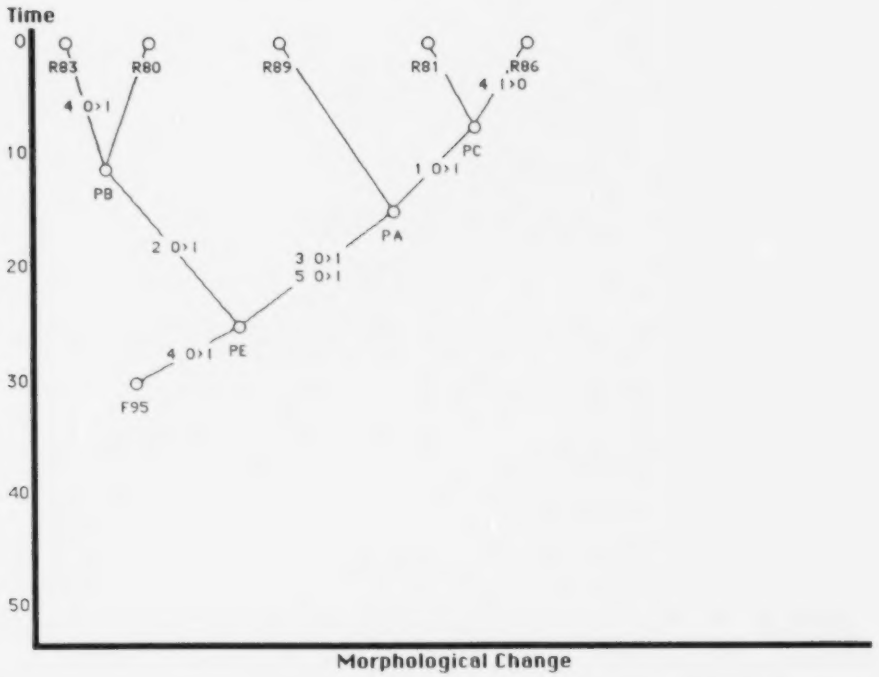

Figure 10. A character optimization of the initial topology of problem 1.4.

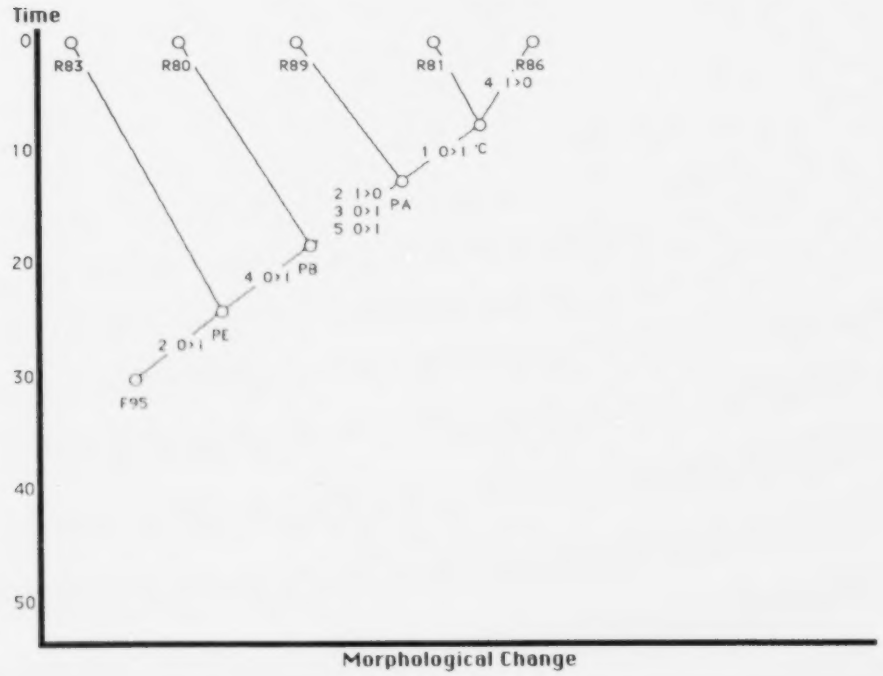

Figure 11. A second character optimization of the third topology for problem 1.4.

Conclusions

This model was initially developed for use in a course that taught domain-specific problem solving using a cognitive apprenticeship approach. Using this approach, the instructor demonstrates problem solving (modeling), helps students solve problems (coaching), and encourages students to solve problems autonomously (fading), until students have developed competence at solving the problems independently. The model can be useful to students to help them interpret the instructor's behavior and to guide to their own problem-solving; by itself the model is insufficient to teach effective problem-solving to students.

These simple tree construction problems allow students to become familiar with the processes used by scientists to explain evolutionary history. All of the experts in the study agreed that the problems were a realistic characterization of the concepts and processes central to their discipline. At the same time, it should be recognized that the processes as presented in this study have been decontextualized and that students, especially the introductory students who might benefit most from solving these problems, should also work with problems constructed from rich data sets including real or realistic imaginary organisms, such as the Caminalcules (Sokal, 1983) or Dendrogrammaceae (Duncan, Philips, & Wagner, 1980).

Students with a tree-based conception of phylogenetic biology should be better prepared to understand evolutionary biology and its central role in the rest of biological theory. The ability to see evolution as a branching and historical structure, rather than a ladder or straight line, lies at the heart of much of modern biology. Discarding the ladder-based approach to conceptualizing evolutionary progress may also help students free themselves from the mythos that some organisms are higher or lower than others. This concept, central to understanding the revolutionary power of Darwin's work, is still elusive to many students. Developing a solid foundation of phylogenetic concepts may

transform the way many students experience these ideas and help foster a less anthropocentric view of the history of life.

Literature Cited

Brooks, D. R., & McLennan, D. A. (1991). Phylogeny, Ecology, and Behavior. Chicago: University of Chicago Press.

Duncan, T., Phillips, R. B., & W. H. Wagner, J. (1980). A comparison of branching diagrams derived by various phenetic and cladistic methods. Systematic Botany, 5(3), 264-293.

Eldredge, N., & Cracraft, J. (1980). Phylogenetic Patterns and the Evolutionary Process: Method and theory in comparative biology. New York: Columbia University Press.

Ericsson, K. A., & Simon, H. A. (1993). Protocol Analysis: Verbal reports as data. Cambridge: MIT Press.

Ericsson, K. A., & Smith, J. (1991). Empirical studies of expertise: Prospects and limits. In K. A. Ericsson & J. Smith (Eds.), Toward a general theory of expertise: Prospects and limits (pp. 1-38). Cambridge: Cambridge University Press.

Jungck, J. & Calley, J. (1985). Strategic simulations and post-Socratic pedagogy: Construction computer software to develop long-term inference through experimental inquiry. American Biology Teacher, 14(2), 137-146.

Larkin, J., & Rainard, B. (1984). A research methodology for studying how people think. Journal of Research in Science Teaching, 21(3), 235-254.

Reif, F. (1983). Understanding and teaching problem solving in physics. In Lectures at the International Summer School on Physics Education, La Londe Les Maures, France:

Ridley, M. (1986). Evolution and Classification: The reformation of cladism. New York: Longman Group Limited.

Sokal, R. R. (1983a). A phylogenetic analysis of the Caminalcules: I. The data base. Systematic Zoology, 32(2), 159-184.

Sokal, R. R. (1983b). A phylogenetic analysis of the Caminalcules: II. Estimating the true cladogram. Systematic Zoology, 32(2), 185-201.

Stewart, J. (1988). Potential learning outcomes from solving genetics problems: A typology of problems. Science Education, 72(2), 237-254.

Stewart, J., & Hafner, R. S. (1994). Research on problem solving: Genetics. In D. L. Gable (Eds.), Handbook of research on science teaching and learning (pp. 284-200). New York: MacMillan Publishing Co.

Wiley, E. O. (1981). Phylogenetics: The principles and practice of phylogenetic systematics. New York: John Wiley & Sons.

Wiley, E. O., Siegal-Causey, D., Brooks, D. R., & Funk, V. A. (1991). The Compleat Cladist (Special Publication No. 19). University of Kansas Museum of Natural History.

Appendix A

A Primer of Phylogenetic Assumptions, Diagrammatic Elements, and Terms

A Primer of Phylogenetic Assumptions, Diagrammatic Elements, and Terms

Assumptions of Phylogenetic Inference:

1. There is only one true phylogeny.
2. Shared characters are the result of homology.
3. The polarity of character states is knowable.

Elements of Phylogenetic Diagrams

Figure 1 illustrates an example phylogenetic tree created using Phylogenetic Investigator. This section describes the phylogenetic tree and its elements. Terms are organized alphabetically at the end with definitions and examples that also reference this tree where possible.

The data matrix from which this diagram is generated appears in the lower right hand corner showing characters in columns and taxa in rows. The intersection between each row and column has a symbol that indicates where that taxon has the apomorphic (1) or plesiomorphic (0) form of the character.

The phylogenetic tree is constructed along two axes. The ordinate represents time divided into 50 units and the abscissa represents morphological change as a continuous, unitless variable. The small circles are nodes. Each node has a designation associated with it. Nodes that begin with "R" represent recent taxa. Nodes that begin with "F" represent fossil taxa. Nodes that begin with "P" are postulated taxa. Lines that link nodes together indicate lines of ancestor/descendant relationship. Some links contain one or more transitions. Each transition (e.g., "1 0>1" or "1 1>0") indicates that the referenced character (1) changed in state either from plesiomorphic to apomorphic (0>1) or reversed from apomorphic to plesiomorphic (1>0) at some point in time along the link on which it appears.

In Figure 1, characters 1-5 are represented as being homologous. Characters 6 and 7 are homoplasious in this diagram. Character 8 is an autapomorphy and is irrelevant to the decision-making process of tree construction. An autapomorphic character is always constructed as a transition immediately prior to the taxon that possesses it.

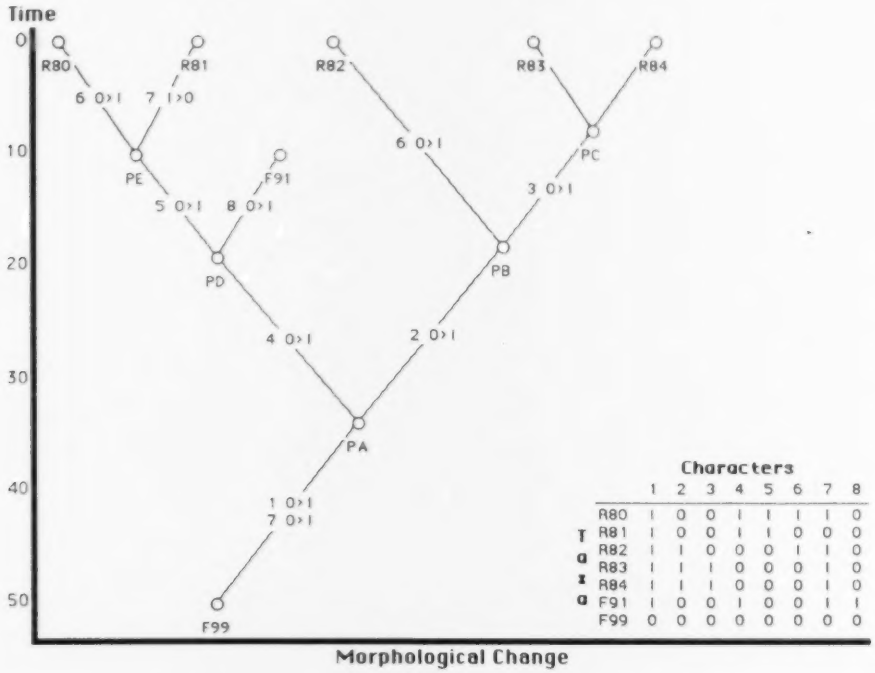

Figure 1. An Example Phylogenetic Tree

Character 1 is a whole-group synapomorphy that supports the existence of postulated ancestor PA. Character 1 is inclusive of all other characters. Character 2, which groups R82, R83 and R84, supports node PB. Character 2 is inclusive of character 3 and exclusive of character 4. Character 3, which groups R83 and R84, supports node PC. Character 4, which groups R80, R81, and F91, supports node PD. Character 4 is inclusive of character 5 and exclusive of character 2. Character 5, which groups R80 and R1, supports node PE.

Character 6 claims that R80 and R82 are a group. For character 6 to be true, characters 2, 5, and 4 would have to be false. In other words, in order to save one step in

character 6, at least three other steps would required. Character 6 is most parsimoniously gained convergently in R80 and R82. Character 7 claims that all of the taxa except for R81 are a group. For character 7 to be true, characters 4 and 5 would have to be false. Saving a step in character 7 would result in at least two added steps Character 7 is most parsimoniously optimized as a reversal in R81.

Terms of Phylogenetic Inference

Ancestor	A taxon, previous in time to a second taxon, from which the second taxon is descended. For example, Figure 1 proposes that a postulated taxon PC is the common ancestor of R83 and R84.
Apomorphy	An evolutionary character, usually coded as "1", that represents an evolutionarily novel state. Character 1 is an apomorphy in all of the taxa of the ingroup (Fig. 1).
Autapomorphy	The transition of a character that is uniquely evolutionarily novel (apomorphic) for a taxon. Character 8 an autapomorphy because it is possessed in the apomorphic state only by taxon F91 (Fig. 1).
Character	A recognizable feature that varies among taxa. For example, among ladybugs, the characters might include the presence or absence of spots. Characters are numbered, polarized, coded, and presented in columns in the data matrix (Fig. 1).
Clade	A monophyletic taxon.
Cladogram	A form of phylogenetic tree that can only show sister-group relationships. Figure 1 illustrates sister-group relationships between all of the taxa, except F99, which is claimed to be a true ancestor of all of the other taxa.
Conflict	A quality of characters that contain incompletely overlapping distributions of apomorphies. Characters 5 and 6 conflict because both are apomorphic for 80, but 5 is apomorphic for 81 and 6 is apomorphic for 82 (Fig. 1).
Convergence	A form of homoplasy whereby two taxa share a character that has appeared independently in separate lineages. Character 6 arises convergently in taxa R80 and R81 (Fig. 1).
Data Matrix	A summary table of states with taxa in rows and characters in columns. The data matrix appears in the lower right-hand corner (Fig. 1).
Descendant	A taxon which is the genealogical product of an earlier taxon. Taxon R84 is a descendant of PC (Fig. 1).

Exclusive	Characters whose distributions of apomorphies do not overlap. Characters 2 and 4 are exclusive of one another (Fig. 1).
Homology	The quality of characters that are shared as the result of common ancestry. See assumption 2. Characters 1, 2, 3, 4, 5 are assumed to be homologous (Fig. 1).
Homoplasy	Characters that are shared due to causes other than homology (evolutionary convergence or reversal). Character 6 is homoplasious and explained using convergence and character 7 is homoplasious and explained using reversal (Fig. 1).
Inclusive	When one character's distribution of apomorphies is a superset of another character's distribution of apomorphies. Character 2 is inclusive of character 3 (Fig. 1).
Ingroup	The group of taxa currently being studied using phylogenetic inference. Taxa R80, R81, R82, R83, R84 and F91 are members of the ingroup (Fig. 1).
Link	A line in between nodes in Phylogenetic Investigator that represents lines of ancestor/descendant relationships. The link between R83 and PC represents a hypothetical ancestor/descendant relationship between R83 and PC (Fig. 1).
Monophyletic	A taxon that includes only the complete set of descendant taxa of an ancestral species. The group of R83 and R84 (and PC) is a monophyletic taxon (Fig. 1).
Node	A circle in Phylogenetic Investigator used to represent a taxon. R80 is a node that represents a taxon (Fig. 1).
Optimization	The process or product of distributing a homoplasious character on a phylogenetic tree. Characters 6 and 7 are optimized in Figure 1.
Outgroup	A group of taxa used to polarize the character states.
Parallelism	A convergence.
Paraphyletic	A grouping of taxa that does not reflect the underlying evolutionary relationships by removing taxa from a monophyletic taxon. A grouping of R82 and R84 is paraphyletic (Fig. 1).
Parsimony	A principle used to justify selecting the hypothesis that requires the fewest transitions and a corollary to assumption 2: By assuming homology, one also selects the hypothesis that minimizes the number of assumptions of homoplasy. The phylogenetic tree in Figure 1 is the most parsimonious explanation of the data.
Phylogenetic tree	A branching diagram that can illustrate both sister group and ancestor/descendant relationships among a set of taxa. Figure 1 is a phylogenetic tree.

Phylogeny	The set of ancestor/descendant relationships that form the genealogy of a set of taxa. A phylogenetic tree (Fig. 1) is a hypothetical representation of these relationships.
Plesiomorphy	A form of a character (state) which is evolutionarily preexisting for the group of taxa under study (the ingroup). Character 2 is retained in the plesiomorphic state by R80, R81, and F91 (Fig. 1). Character 7 occurs in the plesiomorphic state in taxon R81 and this is explained using a hypothesis of reversal (Fig. 1).
Polarity	Whether a form of a character (a state) is considered apomorphic (evolutionary novel) or plesiomorphic (evolutionarily preexisting). This is usually done through comparison with an outgroup.
Polyphyletic	A grouping of taxa that does not reflect the underlying evolutionary relationships by adding unrelated taxa to a monophyletic taxon. A grouping of R81, R83, and R84 would be polyphyletic (Fig. 1)
Reversal	The transition of a character that is apomorphic in some ancestor, changes polarity back to the plesiomorphic state resulting in descendant taxa which are plesiomorphic for that character. Character 7 is optimized as a reversal in taxon R81 (Fig. 1).
Sister group	The most closely related taxon to another taxon. R82 is the sister group to the taxon of R83 and R84 (Fig. 1)
State	A form of a character that is polarized as either apomorphic or plesiomorphic and coded as "1" or "0". For example, among ladybugs, the absence of spots might represent the plesiomorphic state and the presence of spots might represent the apomorphic state.
Steps	The number of transitions required to explain a character or characters. Character 6 is explained in two steps (Fig. 1).
Synapomorphy	The transition of a character that is homologously shared in the evolutionary novel (apomorphic) condition. Character 1 is a synapomorphy for the whole ingroup (Fig. 1).
Taxon	A group of organisms that is given a name. The complete set of taxa descended from a common ancestor is a monophyletic taxon. Incomplete sets are paraphyletic and sets with extra unrelated taxa are polyphyletic. R80, R81, and F91 are a monophyletic taxon because they all are hypothesized to have descended from PD (Fig. 1).
Topology	An arrangement of sister-group or ancestor/descendant relationships among a group of taxa. Figure 1 has only one most parsimonious topology—any rearrangement of the relationships among the taxa would require more steps than the current tree to explain all of the characters.

Transition	A point in time in a lineage at which a character is hypothesized to have changed in state. At some point between 20 and 35 units of time before the present, character 4 is hypothesized to have changed in state in taxon PD (Fig. 1).
Treelength	The steps, or number of transitions, required to explain the data matrix using a phylogenetic tree. Figure 1 requires a treelength of 10 steps to most parsimoniously explain the data in the matrix.